

Supplementary Information for:

Fetal hemoglobin persistence and neurodevelopmental alterations due to *BCL11A* deletions

Anindita Basak[§], Miroslava Hancarova[§], Jacob C. Ulirsch[§], Tugce B. Balci[§], Marie Trkova, Michal Pelisek, Marketa Vlckova, Katerina Muzikova, Jaroslav Cermak, Jan Trka, David A. Dymant, Stuart H. Orkin, Mark J. Daly, Zdenek Sedlacek*, Vijay G. Sankaran*

[§]Contributed equally

*Address correspondence to:

Vijay G. Sankaran, Boston Children's Hospital, 3 Blackfan Circle, CLS 03001, Boston, MA 02115; Email: sankaran@broadinstitute.org

Or to: Zdenek Sedlacek, Charles University 2nd Faculty of Medicine and University Hospital Motol, Plzenska 130/221, 15000 Prague 5, Czech Republic; Email: zdenek.sedlacek@lfmotol.cuni.cz

SUPPLEMENTARY METHODS

Deletion mapping and genetic analysis.

The DNA of the patients was analyzed using microarray comparative genomic hybridization (array CGH) using Human CytoSNP-12 BeadChips (Illumina) according to the manufacturer's protocol. Deletions were confirmed using fluorescence in situ hybridization. The *de novo* inheritance of these deletions was demonstrated by performing similar analysis as above on parental DNA. For analysis of loss-of-function variants in *PAPOLG* and *BCL11A* from the general population, data from the Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>) was used and statistics were calculated as discussed for a population of 6,503 total individuals.

Globin and *BCL11A* locus sequencing.

Sanger sequencing of the human *HBG1/HBG2/HBB* gene locus was carried out to assess for mutations in this locus that could result in elevated HbF levels in the patients (1, 2). The regions where full coverage sequencing was performed included (hg19 coordinates): chr11:5,246,635-5,247,070 and 5,247,668-5,248,372 (*HBB*), chr11:5,270,550-5,271,727 (*HBG1*), and chr11:5,275,474-5,277,265 (*HBG2*). The following HbF associated SNPs in *BCL11A* were genotyped using Sanger sequencing: rs4671393, rs1427407, and rs7606173.

Hematologic and immunologic phenotyping.

Blood samples were obtained from the patients and were subjected to standard clinical hematologic and immunologic assays. This included a complete blood count with a white blood cell differential analysis, reticulocyte count, lymphocyte subset analysis, and immunoglobulin subtype analysis.

Mononuclear cell isolation and RNA analysis.

Mononuclear cells were obtained from peripheral blood by isolation of the buffy coat using Ficoll-Paque (GE Life Sciences) from both patients and normal healthy control children. Briefly, approximately 5 ml blood, diluted in RPMI was layered on Ficoll-Paque and centrifuged at 1300 rpm for 30 min. The buffy coat containing the peripheral blood mononuclear cells (PBMC) was collected and washed in PBS. The PBMC pellet obtained was subject to total RNA isolation using RNeasy Plus Mini Kit (Qiagen). Genomic DNA was eliminated by RNase-free DNase I digestion (Qiagen) during the RNA isolation procedure. Isolated total RNA was quantified on a Nanodrop 2000C instrument (Thermo Scientific). First strand cDNA synthesis and reverse transcription was carried out with the iScript cDNA synthesis Kit (BioRad) in a total volume of 20 µl according to manufacturer's instructions. Gene expression was quantified by quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) using a 96-well plate on a CFX96 Real Time System (BioRad) with iQ SYBR Green Supermix (BioRad) as previously described (3). Primers for qRT-PCR were: *BCL11A* (exon 3 - 4) - 5'-GCCTGGGATGAGTGCAGAAT-3' and 5'-ATGCACTGGTGAATGGCTGT-3'; *PAPOLG* - 5'-CACCCTACCTTCCTGCAGA-3' and 5'-GGATTGAAGTCCGCCCCGAG-3'; *GAPDH* - 5'-TGCACCACCAACTGCTTAGC-3' and 5'-GGCATGGACTGTGGTCATGAG-3'; *HBB* - 5'-CTGAGGAGAAGTCTGCCGTTA-3' and 5'-AGCATCAGGAGTGGACAGAT-3'; *HBG1/HBG2* - 5'-TGGATGATCTCAAGGGCAC-3' and 5'-TCAGTGGTATCTGGAGGACA-3'.

Cell culture and lentiviral transduction

293T cells were maintained in 2 mM L-glutamine containing DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S). For lentivirus

production, 293T cells were transfected with the shRNAs in the pLKO.1 construct with the pVSVG and pDelta8.9 vectors using FuGene 6 reagent (Promega), as we have described previously (3, 4). G-CSF mobilized CD34⁺ cells of peripheral blood from donors were obtained by magnetic sorting and frozen after isolation.

Subsequent to thawing, CD34⁺ cells were resuspended in primary cell culture medium and differentiated to the erythroid lineage, using a previously described culture protocol (4, 5). From day 0 – 7, cells were cultured at a density of 10^5 - 10^6 cells per milliliter in IMDM supplemented with 2% human AB plasma, 3% human AB serum, 1% P/S, 3 IU/mL heparin, 10 ug/mL insulin, 200ug/mL holo-transferrin, 1 IU erythropoietin (EPO), 10 ng/mL stem cell factor (SCF), and 1 ng/mL IL-3. From day 7 on, IL-3 was omitted from the medium. Lentiviral infection occurred on day 2 with subsequent puromycin selection and further differentiation, as previously described (3, 4). For RNA analysis, cells were harvested on day 9 of culture when they were at the basophilic to polychromatophilic erythroblast stages of differentiation.

Transcriptome analysis.

Analysis of RNA sequencing (RNA-seq) data from differentiating human erythroid cells was obtained from the Gene Expression Omnibus accession GSE53983 and analyzed using the Tuxedo suite of tools (6, 7). Single-end reads were aligned to the human genome (build hg19) and transcriptome, both obtained from UCSC genome browser, using Tophat version 2.0.10 and allowing for novel junctions. For representation in a genome browser, aligned replicates were combined and normalized to reads per million. RNA-seq data from developing and adult post-mortem human brain regions was obtained and analyzed, as described (8, 9). This included a total of 524 RNA-seq data

samples that were combined according to age of specimens, and all brain regions were averaged for each age or analyzed separately.

Schizophrenia genome-wide association study (GWAS) reanalysis

We obtained data from a large meta-analysis of 36,989 schizophrenia cases and 113,075 controls (10). In the original analysis, 128 linkage-disequilibrium-independent SNPs exceeded genome-wide significance ($P < 5 \times 10^{-8}$). To examine whether there may be SNPs that reach a threshold of significance that is highly significant on a genome-wide scale, but not meeting the conservative threshold set in the initial report, we examined SNPs that reached a threshold of $P < 5 \times 10^{-7}$. In this subsequent analysis, we noted that there was a significant association peak in the intron of *BCL11A* that is reported here. The schizophrenia GWAS data was analyzed and depicted with the assistance of the Ricopili tool (<http://www.broadinstitute.org/mpg/ricopili/>).

SUPPLEMENTARY REFERENCES

1. Sankaran, V.G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J.A., Beggs, A.H., Sieff, C.A., Orkin, S.H., Nathan, D.G., Lander, E.S., et al. 2012. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *The Journal of clinical investigation* 122:2439-2443.
2. Sankaran, V.G., Xu, J., Byron, R., Greisman, H.A., Fisher, C., Weatherall, D.J., Sabath, D.E., Groudine, M., Orkin, S.H., Premawardhena, A., et al. 2011. A functional element necessary for fetal hemoglobin silencing. *The New England journal of medicine* 365:807-814.
3. Sankaran, V.G., Menne, T.F., Scepanovic, D., Vergilio, J.A., Ji, P., Kim, J., Thiru, P., Orkin, S.H., Lander, E.S., and Lodish, H.F. 2011. MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proceedings of the National Academy of Sciences of the United States of America* 108:1519-1524.
4. Ludwig, L.S., Gazda, H.T., Eng, J.C., Eichhorn, S.W., Thiru, P., Ghazvinian, R., George, T.I., Gotlib, J.R., Beggs, A.H., Sieff, C.A., et al. 2014. Altered translation of GATA1 in Diamond-Blackfan anemia. *Nature medicine* 20:748-753.
5. Hu, J., Liu, J., Xue, F., Halverson, G., Reid, M., Guo, A., Chen, L., Raza, A., Galili, N., Jaffray, J., et al. 2013. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood* 121:3246-3253.
6. An, X., Schulz, V.P., Li, J., Wu, K., Liu, J., Xue, F., Hu, J., Mohandas, N., and Gallagher, P.G. 2014. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* 123:3466-3477.
7. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 7:562-578.
8. Shen, E.H., Overly, C.C., and Jones, A.R. 2012. The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends in neurosciences* 35:711-714.
9. Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K., Gerdts, J., Baker, C., Vulto-van Silfhout, A.T., et al. 2014. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 158:263-276.
10. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421-427.

SUPPLEMENTARY TABLE

Supplementary Table 1. Loss of Function Variants in *PAPOLG*

hg19 Position	Prevalence [¶]	Transcript	Function	cDNA alteration	Protein Alteration
chr2:60995628	A1A1=2, A1R=285, RR=5973	NM_022894.3	frameshift	c.271_283del13	p.(T91Kfs*24)
chr2:60995951	A1A1=0, A1R=1, RR=6254	NM_022894.3	frameshift	c.365_366del2	p.(H122Rfs*5)
chr2:61021851	TT=0, TC=1, CC=6494	NM_022894.3	stop gained	c.1996C>T	p.(R666*)

[¶]R corresponds to the wild type and A1 corresponds to the loss of function allele

SUPPLEMENTARY FIGURE LEGENDS

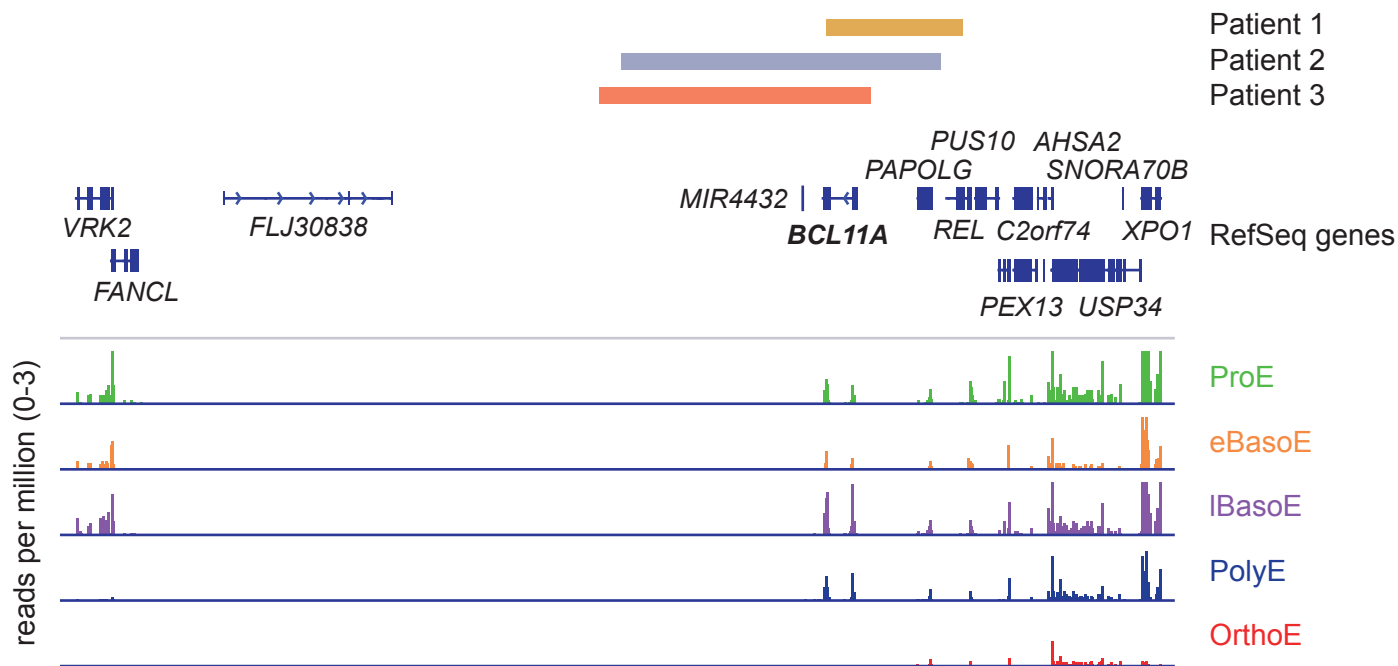
Supplementary Figure 1. Erythroid RNA Expression in the Region of the 2p Microdeletion Syndrome. A depiction of the 2p15-p16.1 region. The position of the patient deletions are shown in orange (Patient 1), blue (Patient 2), and vermillion (Patient 3), and RefSeq genes are shown below. RNA expression is shown below at various stages of human erythroid differentiation. This includes proerythroblasts (ProE), early basophilic erythroblasts (eBasoE), late basophilic erythroblasts (lBasoE), polychromatic erythroblasts (PolyE), and orthochromatic erythroblasts (OrthoE). The height of RNA peaks in each region demonstrates the number of reads per million at that site. This is a zoomed out view of what is shown in Figure 1.

Supplementary Figure 2. *PAPOLG* Does Not Affect Fetal Hemoglobin Levels. (A) The relative expression of *PAPOLG* mRNA in primary human erythroid cells treated with control or *PAPOLG* shRNA containing lentiviruses is shown at day 9 of differentiation. The shRNAs targeting *PAPOLG* (sh1 – 4) are compared with the shRNA targeting luciferase (shLuc) using a Mann-Whitney U test with the p-value as shown. **(B)** The percentage of *HBG1* and *HBG2* in cells with shRNAs targeting controls or *PAPOLG* as in the prior panel shown from day 12 of differentiation (percentages were calculated by measuring *HBG1*/*HBG2* and *HBB* expression and then dividing *HBG1*/*HBG2* by the sum of *HBG1*/*HBG2* and *HBB*). Comparisons are performed using the Mann-Whitney U test as in the prior panel.

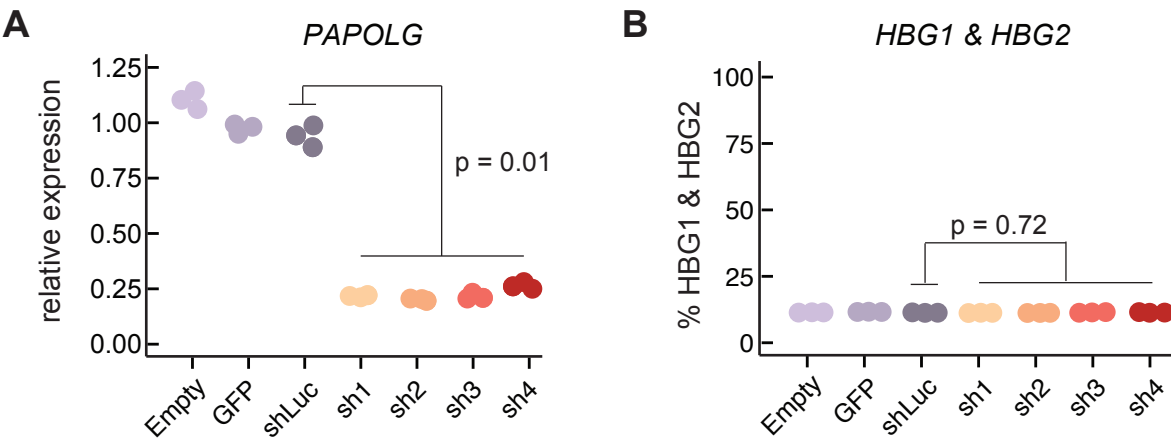
Supplementary Figure 3. Brain RNA Expression of Known Autism Genes. The expression in brain tissue at different developmental stages (from various brain regions that are aggregated here for simplicity) is presented and demonstrates the number of reads per kilobase per million for the genes shown. A locally weighted scatterplot smoothing regression was applied to expression of each gene where each age is an independent and equal time point. Results from this regression are plotted with 95% confidence intervals. All high confidence genes implicated in autism from the Simons Foundation Autism Research Institute are plotted.

Supplementary Figure 4. Spatial Brain Gene Expression During Development. Fine spatial gene expression for *CHD8*, *DYRK1A*, and *BCL11A*, are shown at 15 pcw and 16 pcw. *CHD8* has been previously shown to localize to the intermediate zone, but enrichment was not observed in this region at either time point for *DYRK1A* or *BCL11A*. For simplicity, regions are organized first by lobe and grouped similar to previous investigations of *CHD8*. Abbreviations are as follows: frontal, f; parietal, p; temporal, t; occipital, o; supragenulate nucleus of the thalamus, SG; marginal zone, MZ; outer cortical plate, CPo; inner cortical plate, CPi; SP, subplate zone; intermediate zone, IZ; outer subventricular zone, SZo; inner subventricular zone, SZi; ventricular zone, VZ; frontal polar cortex, fp; dorsolateral prefrontal cortex, dl; dorsomedial frontal cortex, dm-f; ventrolateral prefrontal cortex, vl; orbital frontal cortex, or; posterior frontal cortex (motor cortex), m1; primary somatosensory cortex, s1; dorsomedial parietal cortex, dm-p; posterosuperior (dorsal) parietal cortex, pd; posteroinferior (ventral) parietal cortex, pv; medial temporal-occipital cortex, mt; lateral temporal-occipital cortex, lt; superolateral temporal cortex, sl; inferolateral temporal cortex, il; posterior parahippocampal cortex, ph; midinferior temporal cortex (area 36), t36; caudal midinferior temporal cortex (area TF), tf; midlateral temporal cortex, mlt; primary visual cortex, v1; dorsomedial extrastriate cortex, dm-o; ventromedial extrastriate cortex, vm; midlateral extrastriate cortex, mle.

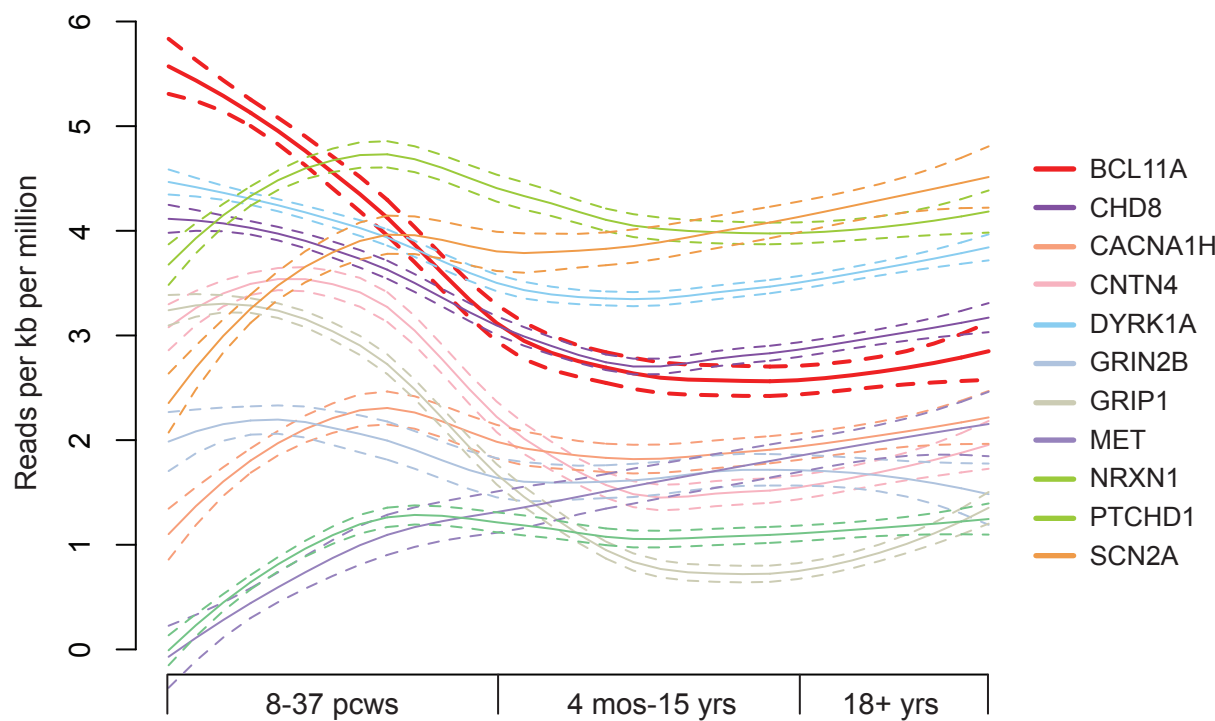
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4

